Session HS2.3.1, Abstract EGU2020-9567
May 6, 2020

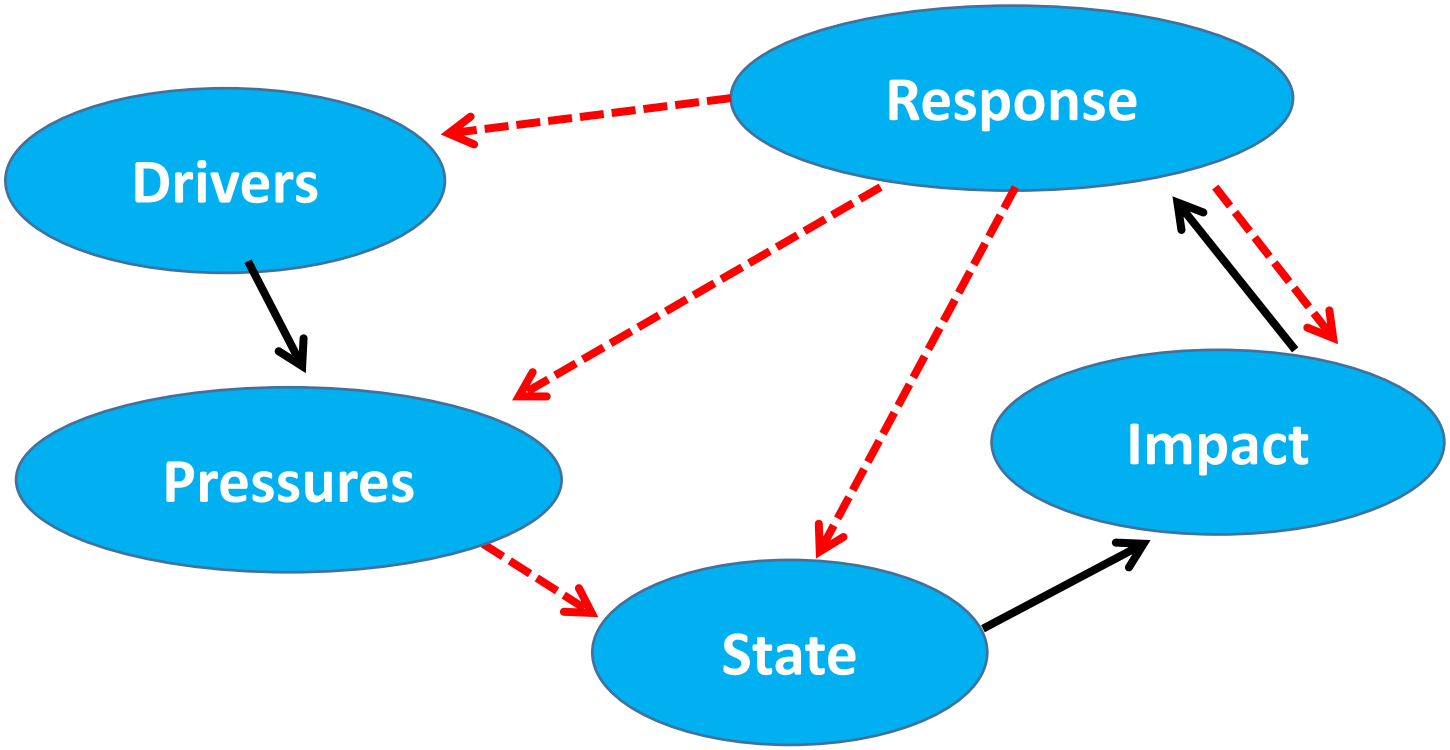
A machine learning model to link ecological response and anthropogenic stressors: a tool for water management in the Tagus River Basin (Spain)

Carlotta Valerio[1,2], Alberto Garrido[2,3], Gonzalo Martinez-Muñoz[4], Lucia De Stefano[1,2]
Email: carval02@ucm.es

[1] : Universidad Complutense de Madrid| [2] : Observatorio del Agua, Fundación Botín | [3] : Universidad Politécnica de Madrid | [4] : Universidad Autónoma de Madrid

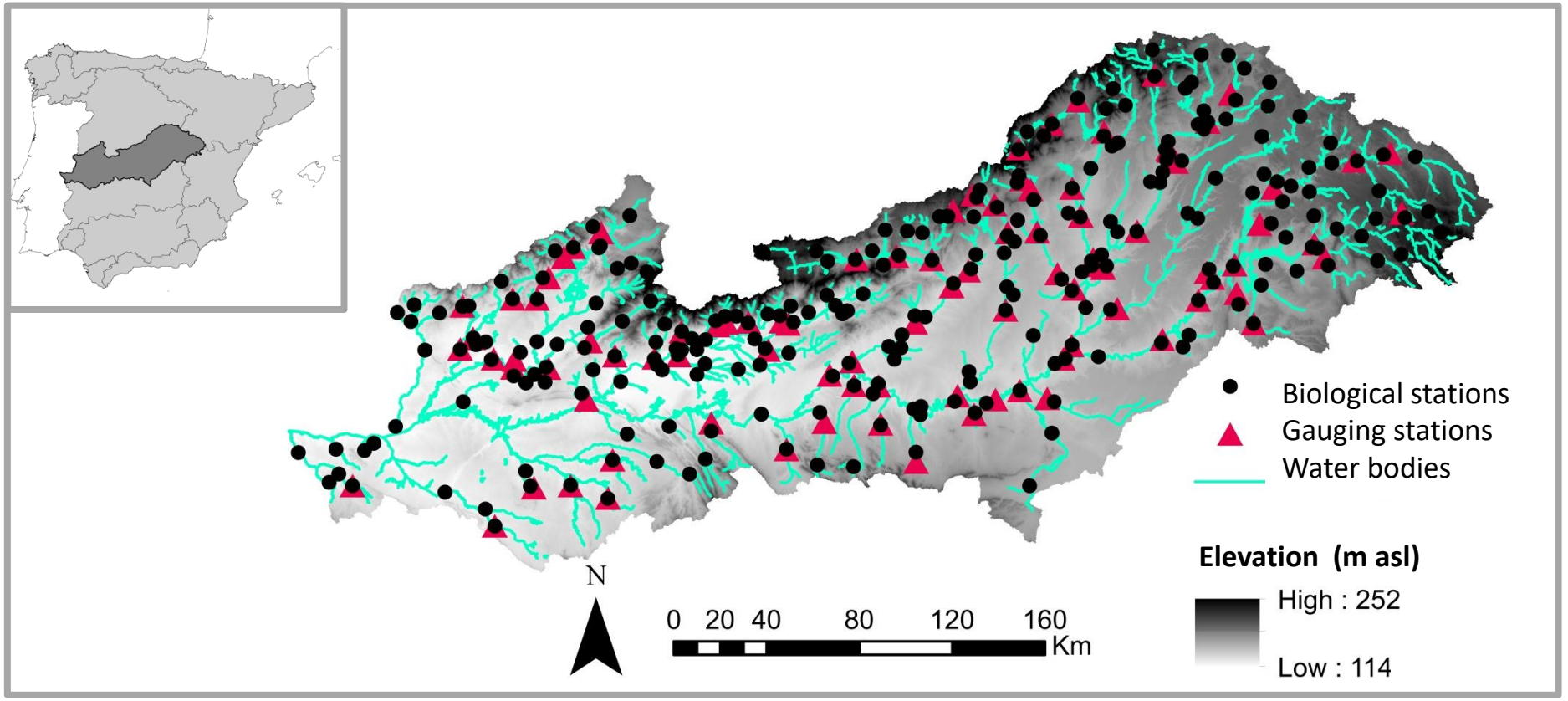
1. RESEARCH CONTEXT

- Water Framework Directive's objective: good ecological status in all the water bodies by 2027.
- Only 40% of European surface waters are in good ecological status.
- The linkage between the pressures acting on the rivers and their ecological status is poorly defined → essential to design effective mitigation measures.

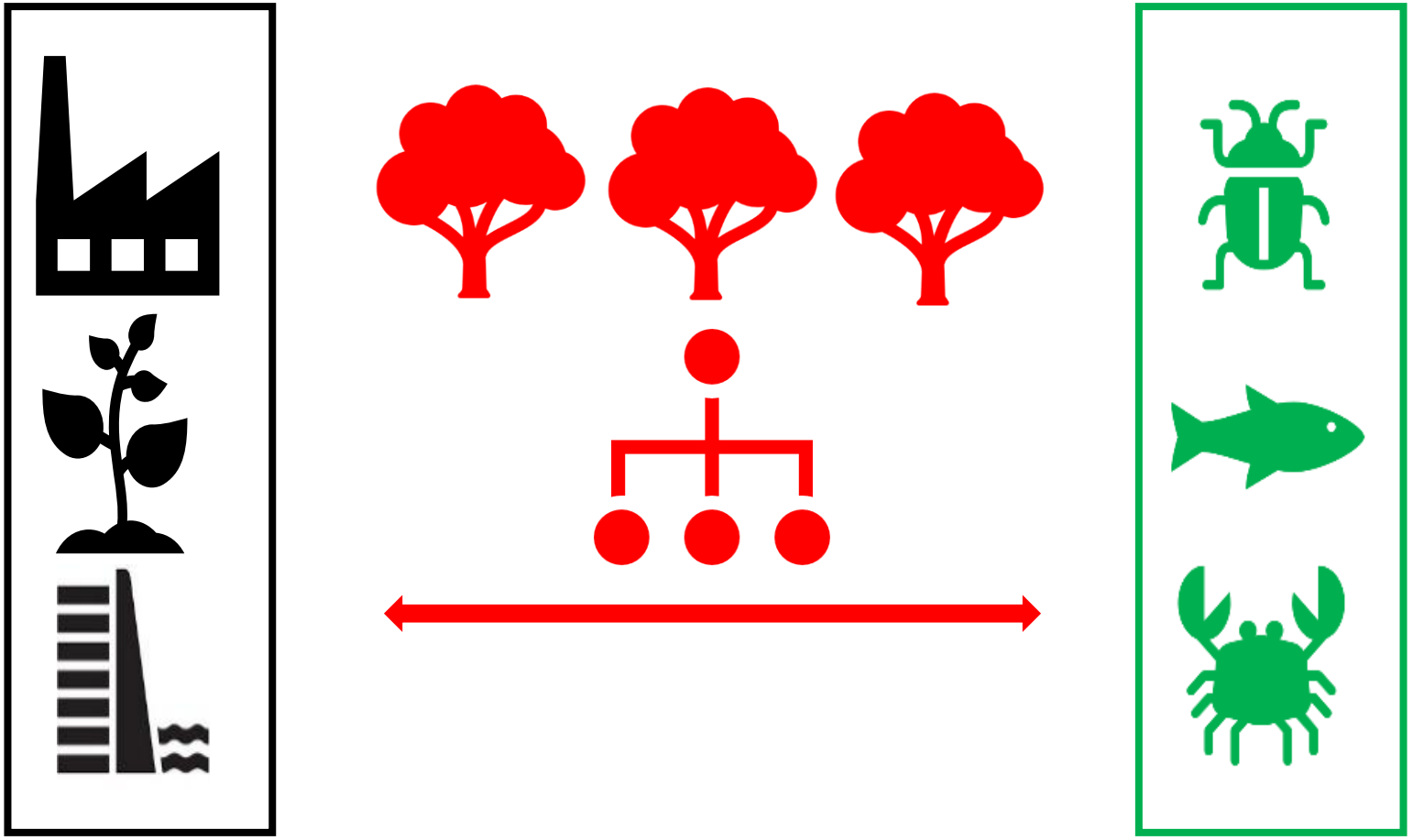


2. RESEARCH AIM

Modelling the relationships between the anthropogenic pressures acting on the system (point and diffuse pollution, alteration of the hydrological regime and hydromorphological degradation) and the biological indices in the Tagus River Basin (Spain) with machine learning techniques.



3. METHODS

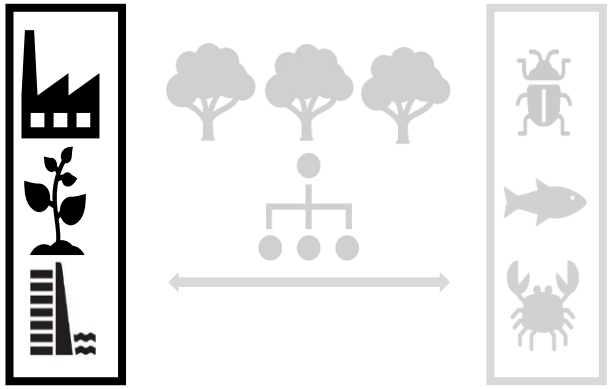


Pressures

Machine learning

Biological indices →
Ecological status

3. METHODS: PRESSURES



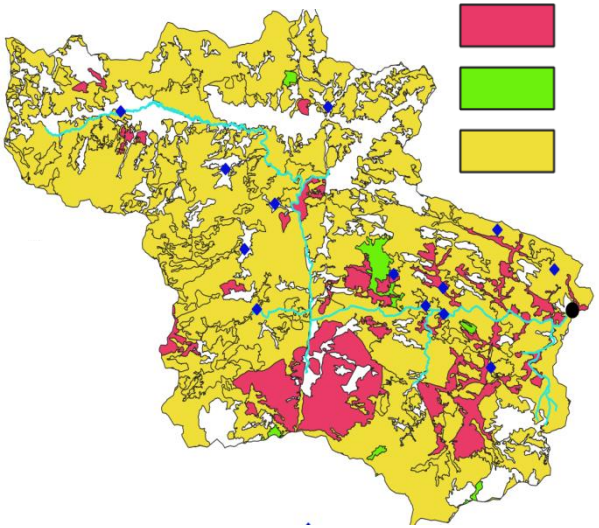
The stressors used as explanatory variables belong to the following categories:

Hydrological regime

IHA indicators: magnitude, frequency, duration, timing

Land use

In the upstream catchment area of each biological sampling station:



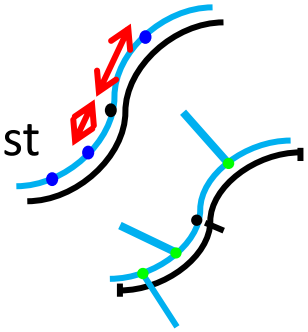
- Urban wastewater discharges
- % Urban areas
- % Forests
- % Agricultural areas

Physicochemical water quality

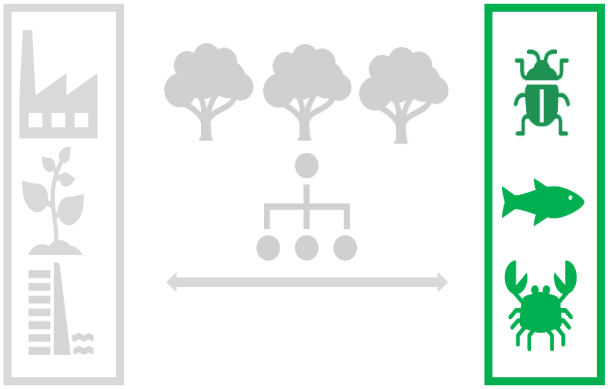
- Conductivity
- pH
- Dissolved oxygen
- Ammonium
- Phosphate
- Nitrate

Hydromorphological indices

a) Alteration of the riparian forest
b) Dam/weir impact indicators
c) Effect of tributaries



3. METHODS: ECOLOGICAL STATUS

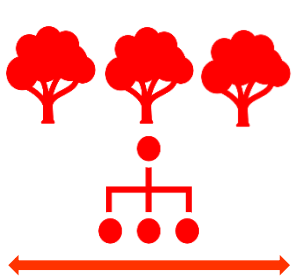


The Biological Quality Elements and the respective biological indices used in the Tagus basin are:

- **Macroinvertebrates** (IBMWP index)
- **Diatoms** (IPS index)

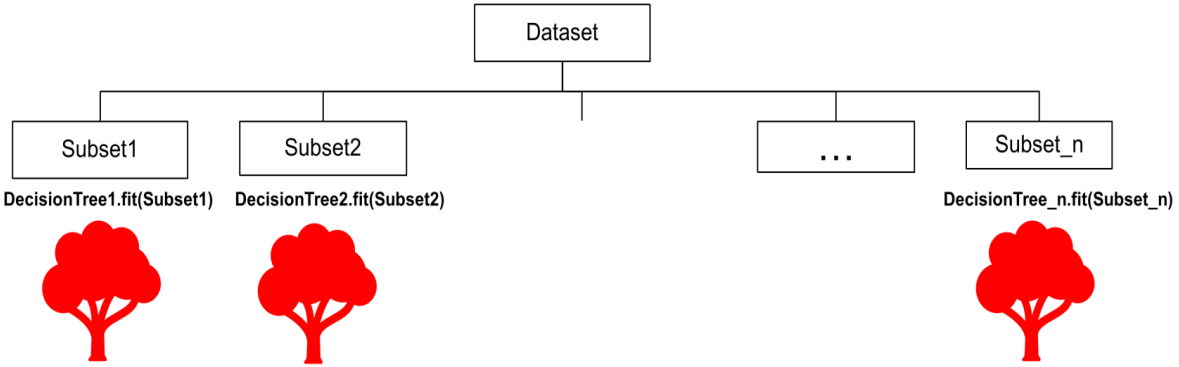
The biological sampling campaigns are usually carried out once a year in spring. The data are available for the 2008-2018 period.

3. METHODS: MACHINE LEARNING MODEL

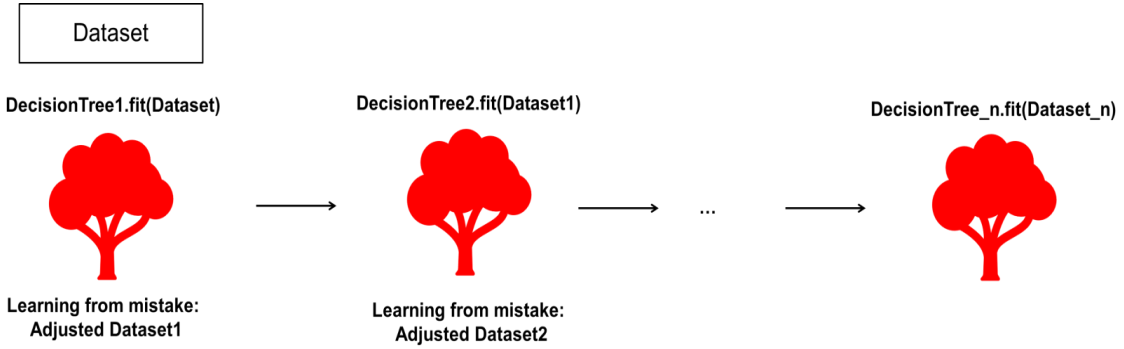


The performance of two machine learning algorithms is compared:

- **Random Forest (RF)**



- **Boosted regression Tree (BRT)**



4. RESULTS

- The modelling quality turns out to be satisfactory.
- RF performs better than BRT.
- Land use categories and nutrient concentrations are ranked as the most significant stressors in the model.
- Once fitted, the model is used to predict the biological indices under different scenarios of nutrient concentrations and riparian habitat quality .